Plasticity of the binding pocket in peptide transporters underpins promiscuous substrate recognition

Vadim Kotov[1,2,8]*, Maxime Killer[1,2,3]*, Katharina E. J. Jungnickel[1,2]*, Jian Lei[1,2,4]*, Giada Finocchio[1,2], Josi Steinke[1,2], Kim Bartels[1,2], Jan Strauss[1,2], Florine Dupeux[5], Anne-Sophie Humm[5], Irina Cornaciu[5,9], José A. Márquez[5], Els Pardon[6,7], Jan Steyaert[6,7] and Christian Löw[1,2#]

[1] Centre for Structural Systems Biology (CSSB), Notkestraße 85, 22607 Hamburg, Germany.
[2] European Molecular Biology Laboratory (EMBL) Hamburg, Notkestraße 85, 22607 Hamburg, Germany.
[3] Collaboration for joint PhD degree between EMBL and Heidelberg University, Faculty of Biosciences.
[4] State Key Laboratory of Biotherapy and Cancer Center, Sichuan University, No. 17, Block 3, Southern Renmin Road, Chengdu, Sichuan 610041, People's Republic of China.
[5] European Molecular Biology Laboratory (EMBL) Grenoble, 71 Avenue des Martyrs CS 90181, 38042 Grenoble Cedex 9, France.
[6] Structural Biology Brussels, Vrije Universiteit Brussel (VUB), Brussels 1050, Belgium.
[7] VIB-VUB Center for Structural Biology, VIB, Brussels 1050, Belgium.
[8] Present address: Evotec SE, Essener Bogen 7, 22419 Hamburg, Germany.
[9] Present address: ALPX s.a.s. 71 Avenue des Martyrs, 38000 Grenoble, France.

*** Authors contributed equally**

[#] Corresponding author
Christian Löw
European Molecular Biology Laboratory Hamburg, Notkestrasse 85, D-22607 Hamburg, Germany.
Phone: +49 40 8998 87570
e-mail: christian.loew@embl-hamburg.de
Twitter handle: @AllUNeedIsLoew

Summary

Proton-coupled oligopeptide transporters (POTs) are promiscuous transporters of the Major Facilitator Superfamily, that constitute the main route of entry for a wide range of dietary peptides and orally administrated peptidomimetic drugs. Given their clinical and pathophysiological relevance, several bacterial and mammalian POT homologs have been extensively studied on a structural and molecular level. However, the molecular basis of recognition and transport of the wide range of peptide substrates has remained elusive. Here we present 14 X-ray structures of the bacterial POT DtpB in complex with chemically diverse di- and tripeptides, providing novel insights into the plasticity of the conserved central binding cavity. We analyzed binding affinities for more than 80 peptides and monitored uptake by a fluorescence-based transport assay. To probe if all natural 8400 di- and tripeptides can bind to DtpB, we employed state-of-the-art molecular docking and machine learning and conclude that peptides of a specific subset with compact hydrophobic residues are the best DtpB binders.

Keywords

major facilitator superfamily; SLC15; peptide transporter; macromolecular crystallography; nanobody; promiscuity; docking; NanoDSF; Rosetta; plasticity

Introduction

Living cells have to adapt rapidly to environmental changes to maintain nutrient homeostasis, which requires the use of different nitrogen-containing nutrients. Hence, cells express a variety of genes to ensure the scavenging of alternative nitrogen sources such as amino acids or peptides [1]. Peptide transporters of the major facilitator superfamily (MFS), namely the proton-dependent oligopeptide transporter (POT) family, provide the cell with valuable nitrogen and carbon sources by mediating the uptake of di- and tripeptides [2]. POT members are known to have an overall helical arrangement termed "MFS-fold", formed by two helical bundles (N- and C-terminal bundle) that are related by a pseudo-two-fold symmetry [3-8] and function according to the alternate access mechanism [9-13]. Additionally, they carry specific sequence signature motifs important for proton coupling, ligand binding, and transport [2,3,14]. They are known to be highly promiscuous, expected to transport almost all 8400 di- and tripeptides composed of proteinogenic amino acids [2].

In *E. coli*, four members of the POT family were identified and termed di- and tripeptide permease (Dtp) A to D [15-17]. Experimental structures of DtpA, DtpC and DtpD were reported recently [7,8,18]. Transport inhibition experiments indicated that the binding sites of DtpA and DtpB interact with a large number of substrates and peptidomimetic drugs in a similar fashion as the extensively studied mammalian homolog PepT1 [16,19-24]. DtpC and DtpD, however, were classified as atypical POTs, as they were shown previously to favour positively charged peptides as substrates [18,22,25-30].

In the last decade, over 50 entries of the POT family were deposited in the Protein Data Bank (PDB), representing eleven bacterial and two mammalian homologs, bound to eight unique natural di- and tripeptides (Ala-Phe, Ala-Glu, Ala-Gln, Ala-Leu, Phe-Ala, Phe-Ala-Gln, Ala-Ala-Ala, alafosfalin), peptidomimetic drugs (valaciclovir, valganciclovir, 5-aminolevulinic acid) and a potent transport inhibitor Lys[Z(NO_2)]-Val [3-8,10,18,31-41]. These efforts revealed the basic principle underlying peptide binding in POTs which can be described as an electrostatic clamping mechanism between the invariable part of peptides (N- and C- termini as well as the peptide backbone) and conserved residues in the transporter (mainly via arginine, lysine, glutamate, asparagine and tyrosine residues). In addition, the molecular changes during transport have been recently described in human PepT1 and PepT2, providing insights into the dynamics of peptide transporters and the local rearrangements of the binding site through-out the full transport cycle [31]. However, it still remains unclear how POTs can recognize and transport such a vast variety of peptides due to the lack of high-resolution structures of POTs bound to chemically diverse substrates.

Here, we determined crystal structures of DtpB complexed with 14 different di- and tripeptides, providing novel insights into the plasticity of the conserved central binding cavity in response to a

wide range of chemically diverse peptides. We thereby also complete the entire family of experimental POT structures from *E. coli* ranging from DtpA to D. Moreover, we measured binding affinities for more than 80 peptides using the thermal shift method and employed a fluorescence-based transport assay to monitor the uptake of peptides into liposomes reconstituted with DtpB. Our analysis indicates that high affinity peptides are only poorly transported and rather act as inhibitors, while peptides in the medium affinity range display the highest transport rates. Finally, we employed state-of-the-art molecular docking and machine learning to probe if all 8400 di- and tripeptides composed of proteinogenic amino acids can bind to DtpB and conclude that a specific subset of peptides with compact hydrophobic residues are the best DtpB binders.

Results and Discussion

Structures of DtpB-peptide complexes stabilized by nanobody 132

In order to obtain highly diffracting crystals of DtpB, we selected conformation-specific nanobodies after immunizing a noninbred llama [42]. Out of 31 recombinantly expressed and purified nanobodies, 14 bound DtpB with a dissociation constant of 30 nM or lower as evident by bilayer interferometry (BLI) measurements (Figure 1, Supplementary Figure 1, Supplementary Tables 1 and 2). They were used as crystallization chaperones in subsequent crystallization trials and nanobody 132 (Nb132) emerged as the most promising binder for co-crystallization approaches. DtpB-Nb132 was initially incubated with the tripeptide Ala-Leu-Ala (ALA), and well-diffracting crystals grew using the vapor diffusion method. The structure of DtpB was determined using the atomic model of DtpA [7] (PDB accession number 6GS4) as search model for molecular replacement. Strong positive peaks in the difference electron density map (i.e. F_{obs}-F_{calc}) indicated the presence of additional electrons/atoms in the periplasmic region of DtpB, and within the central cavity, allowing modelling of Nb132 and the tripeptide.

DtpB crystallized in an inward facing open (IF open) state. In this conformation the central cavity of the transporter is open to the cytoplasm and closed to the periplasm (Figure 1). DtpB adopts the canonical MFS fold, characterized by two six transmembrane helix bundles. The N- and C-terminal bundles are linked by the HA-HB helices, as observed in other bacterial POTs [3-8,10,18,32-35,37,39]. The overall structure of DtpB is similar to DtpA, DtpC, and DtpD with C_α atom RMSD (root-mean-square deviation) values of 0.9, 1.3, and 1.2 Å, respectively. In DtpB, the IF open state is stabilised by electrostatic interactions between the two bundles i.e., Y31-Y285; G35-Q315; Q49-V440; Y64-

Y282; and Y149-E393. Nb132 further stabilizes the closure on the periplasmic side by interacting with the periplasmic surface of the transporter through polar contacts (Figure 1B).

In order to obtain a broader vision of the plasticity of the binding site in response to peptides possessing various chemical groups, the co-crystallization efforts were continued with an in-house library of 82 peptides. Several batches of the DtpB-Nb132 complex were prepared, incubated with a peptide, and then dispensed robotically in 96 well screens containing three sets of crystallization conditions. If the diffraction resolution of an obtained crystal was better than 4 Å, the chemical screens were further refined around the best conditions. In this campaign, X-ray diffraction data were collected and analyzed for more than 2000 crystals with the help of automated crystallography pipelines based on the CrystalDirect technology [43,44]. For each dataset, with a resolution limit higher than 3.5 Å, the presence or absence of a peptide was initially assessed by using an atomic model of DtpB devoid of any substrate, to calculate a difference electron density map (i.e., F_{obs}-F_{calc}). There, strong positive peaks within the central cavity indicated the presence of a ligand. The co-crystallized peptide was then modeled inside the positive peak, as a di- or trialanine moiety first, and then mutated to its original sequence, as the signal improved during the refinement steps. In the final validation, OMIT maps excluding the modeled ligands, were calculated for each structure (Supplementary Figure 2). In summary, we obtained 14 unique peptide bound datasets in a resolution range between 2.0 and 2.8 Å with the following peptides: Ala-Phe (AF), Ala-Ile (AI), Ala-Leu (AL), Ala-Gln (AQ), Ala-Val (AV), Ala-Trp (AW), Lys-Val (KV), Met-Ser (MS), Asn-Val (NV), Ser-Leu (SL), Ala-Leu-Ala (ALA), Ala-Phe-Ala (AFA), Ala-Pro-Phe (APF) and Ala-Trp-Ala (AWA) (Figure 2, Supplementary Figure 3, Table 1). This represents a large portfolio of di- and tripeptides for POTs compared to the present literature data. These structures now allowed us to analyze changes and local rearrangements in the binding site of DtpB and shed light on how promiscuity is achieved in this transporter family.

Plasticity in the peptide binding pocket to accommodate diverse peptides

All DtpB complexes crystallized in the same space group with very similar cell dimensions (Table 1). A superposition of the structures highlights that the N-termini of all di- and tripeptides are anchored in a similar manner (Figure 2). The primary amine is steadily hooked by N153, N318, and E393 of DtpB. This triad of residues remains fixed in all structures, and is conserved in all prototypical POTs, with the exception of the peptide-histidine transporter 1 (PHT1), where N318 is replaced by an aspartate. Additional residues of the binding site form a pocket around the N-terminal residue of the substrate (Supplementary Figures 3 and 4). These include Y31, Q34, S156, S156, L160, M288, and P319. Together with the conserved N153, N318, E393 triad, they constitute the so-called P1 pocket. The chemical diversity of N-terminal residues in the co-crystallized tripeptide data sets

was poor (only alanine residues), but richer for dipeptides (i.e., alanine, lysine, serine, methionine, asparagine). P1 does not undergo conformational changes in presence of these various residues, however, polar interactions stabilize certain substrates (Supplementary Figure 3, Figure 2). Notably, Q34 interacts with the ε-amino group of K1* in KV; N318 with the thioester group of M1* in MS; S156 and N318 with the hydroxymethyl group of S1* in SL; and S156, and N318 with the carboxamide group of N1* in NV. The C-terminus of dipeptides adopted a constant position, stabilized by R27 and occasionally by K123 as well.

The picture is different for the second residue of the substrate. To adapt the central binding cavity to the different sizes of side chains carried by the C-terminal residues of dipeptides (P2 pocket), two rotamer conformations of Y64 and Q424 are possible (Supplementary Figure 5). They control the volume of the 'upper region' of the P2 pocket. For instance, in presence of the small dipeptide AV, the upper region of P2 is tightened by the conformation of Y64 and Q424, while it is widened in the presence of the bulky dipeptide AW (Supplementary Figure 5). Additional residues such as Y285 and F428 further fine-tune the upper region of P2. Of note, Y285 and Y64 delimit the P2 pocket from the P1 and P3 pockets respectively (Figure 2). Unlike other side-chains fitting in P2, the indole moiety of W2* in the tripeptide AWA extends further down, towards the cytosolic side of the C-bundle (Supplementary Figure 5). This 'lower region' of P2 (i.e. L401, W420, and F421) is more flexible, and closes up to stabilize W2*. Polar interactions also occur between P2 and the peptides. Y282 stabilizes the indole ring of W2* in AWA as well as the carboxamide group of Q2* in AQ (Supplementary Figure 3 F,J). In addition, the hydrophobic side chains in the second position (i.e. in the peptides AV, AL, AI, AF, AW, ALA, AFA, AWA) are increasingly stabilized as a function of their size, and through contraction of the upper region of P2.

The backbone coordination of tripeptides withstands different torsion angles around amide bonds. Since the primary amine of the N-terminus remains hooked in place between N153, N318, and E393, the carboxylic group of the C-terminus is subsequently shifted or rotated, resulting in different poses (Supplementary Figure 3, Figure 2). For instance, the carboxylic group of AWA coincides with the ones of dipeptides. This co-localization is achieved by a kinked backbone geometry of the tripeptide. AFA and ALA are not kinked, but stretched. In comparison, in dipeptides AL and AF a conformational change in K123 creates sufficient space for the carboxylic group of L2* and A3*, to fit in a position preserving the stabilizing salt bridges with R27 and K123 (Supplementary Figures 3 and 6). Finally, in the case of the tripeptide APF, the proline residue restrains the backbone and evicts the C-terminus from the positively charged patch formed by R27 and K123, and the bulky phenyl group of F3* extends towards the cytosolic side of the N-bundle, in P3 (Supplementary Figures 3 and 6). The overall plasticity of the binding site of DtpB is illustrated in Movie 1.

In summary, the N153, N318, E393 triad is a common anchor point of peptides N-termini, in agreement with earlier literature [32,45,46]. We find that the C-termini of peptides are often stabilized by R27 and K123, but the latter is not mandatory, contrary to the previously suggested model [14]. Recent molecular dynamics (MD) studies using human PepT2 and the tripeptide AAA [36] support this observation and suggest that peptides engage with the binding pocket via A1* first, before being tightly locked in place by the triad (N192, N348, E622 in human PepT2). The simulations also indicate that R27 and K64 later contribute to further stabilization of the C-terminus. Importantly, K64 (Q34 in DtpB) was not essential in DtpB to coordinate the tripeptide APF and K64 is generally not conserved among POT homologs. It is likely that the presence of bulky side chains in the N-terminal position, would cause local changes in P1, but we did not succeed in determining structures of such complexes. The versatility of P2 was previously described with rearrangements of Y68 and W427 in a POT from *Streptococcus thermophilus* ($PepT_{St}$) [32,33], which correspond to Y64 and W420 in DtpB. Here, these two residues critically contribute to adapt P2 to the various co-crystallized peptides, but other residues (Y282, Y285, L401, Q424, F428) are also involved. Except for the different rotamer conformations of K123 to enable various positions of tripeptides, the P3 pocket was rather stable compared to P2. Although these observations indicate that the plasticity of POTs originates mainly in the P2 pocket, many more combinations of peptides are able to bind to and be transported by DtpB and other prototypical POTs. Besides, not all residues involved in ligand promiscuity are conserved in POTs. This could explain the differences of substrate affinities between homologs reported in the literature.

Determination of peptide affinities using a thermal unfolding assay

To shed further light on the interactions of different di- and tripeptides with DtpB and determine binding affinities for a large set of peptides, we employed a thermal unfolding assay. This approach is commonly used to characterize the stability of proteins and their functional interactions [47-50]. In particular, the stability of proteins (as judged by the melting temperature T_m) increases in a concentration-dependent manner when a ligand is added (Figure 3A). Various concepts were proposed to obtain a ligand dissociation constant (K_D) based on the ligand-induced shifts of T_m [51,52]. These approaches typically assume classic thermodynamic behavior of proteins during unfolding, i.e. the equilibrium is quickly reached at all temperatures, and protein unfolding is fully reversible. The latter is rarely observed in practice, so a kinetic description of the unfolding process should be used instead [53,54]. Recently Hall introduced and validated a kinetic model to determine affinities from thermal shifts (Figure 3B) [55]. We modified Hall's model to take advantage of modern non-linear curve

fitting methods (see Materials & Methods) and applied it to DtpB titrated with di- and tripeptides. Titration curves could be fit with high confidence for peptides in a broad affinity range (Figure 3C). The resulting K_D was not strongly affected by assay conditions [56] (Supplementary Figure 7A). Importantly, our approach allows us to predict K_D of new peptides obtained from T_m measurements at a single concentration (see Materials & Methods, Supplementary Figure 7B). The rank order of the K_D agreed well with an orthogonal technique (microscale thermophoresis (=MST), Supplementary Figure 7C). With this approach we could efficiently measure K_D values for 82 di- and tripeptides (Figure 3D) to DtpB and cover a broad region of the peptide chemical space. As expected, determined affinities span several orders of magnitude, with the tightest binders showing affinities in the low µM range while others interacted with DtpB only poorly or not at all.

Tight peptide binding is not associated with transport

The core biological function of a transporter is its ability to move molecules across the lipid bilayer. To gain more detailed insights into the uptake of various peptides via DtpB, we established a robust transport assay in liposomes, termed 'pyranine assay'. Since the peptide uptake by POTs is coupled to protons [57-59] we used the pH sensitive fluorescent dye pyranine [60] to indirectly monitor peptide transport into DtpB-containing liposomes. Such an approach was recently applied to characterize the bacterial POT $PepT_{St}$ [61]. We utilized similar experimental conditions to follow transport activity of DtpB reconstituted into liposomes (Figure 4A). To validate the assay, we initially confirmed transport of the known POT substrates Ala-Ala (AA) and Gly-Gly (GG) by DtpB (Figure 4B) and related transporter $PepT_{St}$ [16,61] (data not shown).

To account for batch-to-batch variations between different liposome reconstitutions, we developed a non-linear curve fitting procedure to quantify the obtained transport curves in this assay (see Materials & Methods). In brief, the experimental data were corrected for the empty liposome signal and then fit to a single exponential decay function to obtain the time constant τ (tau) and the amplitude of the transport curve. The ratio of the amplitude to τ corresponds to the initial transport rate at time-point t_0 (slope(t_0)), corresponding to the time of the addition of valinomycin (Figure 4B). Transport measurements of the substrate AA and buffer alone served as a positive and a negative control and were used to normalize slope(t_0) for all measured peptides to range from 0 to 1 thus obtaining slope$(t_0)_{rel}$. Quantification of transport rates at varying substrate concentrations allowed us to determine the apparent Michaelis-Menten constants (K_M) for the dipeptides AA (0.29 ± 0.04 mM) and GG (2.63 ± 0.65 mM (Supplementary Figure 8A). These values are within the expected range for POTs [10,40]. In addition, we used surface electrogenic event reader ($SURFE^2R$) [62] as an orthogonal

technique to verify the uptake of GG and AA by DtpB. In this assay an electrochemical gradient is absent, and the transport of peptides is driven by an excess of substrate in the external buffer. With this, we could confirm that AA and GG are transported by DtpB, however, the apparent K_M values were 20-30 fold higher compared to the electrochemical gradient driven conditions as used in the pyranine assay (Supplementary Figure 8).

Next, we determined slope$(t_0)_{rel}$ for 24 di- and tri-peptides using the pyranine assay, covering a broad spectrum of binding affinities as determined by thermal unfolding (Figure 4C,D). In the context of available K_D values, slope$(t_0)_{rel}$ forms a bell-shaped distribution (Figure 4D). Peptides that poorly bound to DtpB or not at all in our thermal shift assay exhibited low or no transport. This indicates that peptides with very low binding affinities would not initiate the transport cycle since they do not reside long or well enough bound in the binding site for any conformational changes of the transporter to occur. Alternatively, the correlation of low binding affinities and no apparent transport might also result from clashes or unfavoured positioning of the peptide in the binding site. The highest slope$(t_0)_{rel}$, i.e. peptides with at least 20% of the AA transport rate, could be observed for peptides with K_D values in the range of ~100 µM to ~2.5 mM. To confirm this observation, we measured transport for three more peptides picked from this affinity range: AH (K_D = 0.10 ± 0.04 mM), MT (K_D = 0.57 ± 0.07 mM) and AK (K_D = 1.77 ± 0.09 mM) (Figure 4D, red dots). Of those, only MT was transported suggesting that a K_D in a specific range is required, but not the only determining factor to enable transport by DtpB.

Our findings demonstrate that tightly bound peptides are either transported very slowly or act as inhibitors of transport. Interestingly, previous characterization of $PepT_{St}$ using the pyranine assay [61] demonstrated transport of peptides AF and ALA, which are not transported by DtpB. To confirm the transport inhibitory effects of tightly binding peptides, we measured uptake of AA in presence of increasing concentrations of the high affinity binder AF (Figure 4E). The obtained IC_{50} value for AF is 2.37 ± 0.68 mM, which corresponds to an inhibitory constant K_i of 0.25 ± 0.08 mM. The IC_{50} value is two orders higher than the previously reported IC_{50} of ~ 0.027 mM for competition between radiolabeled AA and AF for $PepT_{St}$ [10]. This difference can be attributed to the different peptide concentrations used in the assay (2.5 mM of AA used in this work versus radiolabeled AA used at 0.03 mM concentration [10]). We note that with our assay setup we did not aim to determine the number of protons being transported along with each peptide, though it was previously shown for $PepT_{St}$ that this number may vary between di- and tripeptides [61]. Consequently, the amplitude of the signal in the pyranine assay will be higher for peptides that carry more protons, while the time constant τ is not affected.

DtpB preferentially binds small hydrophobic peptides

A more detailed analysis of the peptide K_D dataset (Figure 3G) indicates that DtpB preferentially binds peptides that are hydrophobic with a molecular weight below 300 Da (Supplementary Figure 9A). Since our dataset accounts for less than 1% of all possible di- and tripeptides made from proteinogenic amino acids (n = 8400), we asked whether the available crystal structures of DtpB-peptide complexes can be used to formulate more precise peptide recognition rules. For this, we performed flexible docking all of possible di- and tripeptides into DtpB by the Rosetta FlexPepDock protocol [63] and used a modified hit selection procedure (see Materials & Methods). This allowed us to predict the placement of the peptide inside DtpB for all 8400 di- and tripeptides (Figure 5A, Supplementary Figure 10).

The Rosetta energy function is expressed in energy units (kcal/mol) and was recently calibrated to reliably represent real energies observed in protein molecules [64]. On the other hand, it is impossible to account for all contributing factors that influence ligand binding without a thorough MD simulation. Thus, the scores reported by Rosetta or other docking applications generally show low or no correlation with experimentally observed affinity [65,66]. Indeed, the rank order correlation (Spearman ρ) between the Rosetta score for docked peptides and their experimentally measured K_D was only -0.28 (data not shown). Therefore, we performed multiple linear regression by using individual Rosetta energy terms (e.g. electrostatic interactions, optimal placement of rotamers [64]) as independent variables (features) and negative decimal logarithm of K_D as the dependent variable. The obtained models performed poorly in cross-validation (CV) tests with coefficient of determination (R^2 score) close to 0 or negative (data not shown). Finally, we separated the di- and tripeptides with experimentally determined K_D into two classes: binders ($K_D \leq 1$ mM) and non-binders ($K_D > 1$ mM) and trained a logistic regression classifier to distinguish these classes based on Rosetta energy terms (Figure 5B). Average area under curve (AUC) of the receiver-operator characteristic (ROC) during CV of this classifier was 0.67 ± 0.04 (see Materials & Methods), suggesting that it outperforms random classification (AUC 0.5). AUC of the classifier trained with full data was 0.88. The particular value of logistic regression is that it provides a probability estimate [67] for each sample to belong to each class (termed 'binder probability' in this work). If the peptide's predicted binder probability is less than 0.4, then it is very likely to have an experimental K_D value above 1 mM (Figure 5B). On the other hand, several peptides with binder probability above 0.6 (AE, LA, LGG, LYA, MAS, RF) still may be poor binders when tested experimentally. This can be attributed to the fact that the solvent interaction is not modeled, which for instance plays an important role in peptide recognition by periplasmic binding proteins OppA [68] and DppA [69]. Furthermore, 'structural snapshots' (as opposed to an MD simulation) are fundamentally limited in capturing the entropic component of binding [62].

Still, the proposed classification approach allows us to exclude obvious non-binders (binder probability below 0.4) which constitute 4192 out of 8400 peptides (Supplementary Figure 9B). Only 727 peptides have binder probability over 0.99, i.e. in the search of new peptide binders to DtpB it is sufficient to experimentally check ~ 9% of all possible di- and tripeptides.

Next, we used the binder probability as a proxy to explore the peptide recognition landscape of DtpB (Figure 5C,D, Supplementary Figure 11). We observe critical importance of the first position in binding for both di- and tripeptides (Figure 5C, Supplementary Figure 11). In particular, polar amino acids but also bulky hydrophobic amino acids tend to decrease the binder probability (Figure 5C, Supplementary Figure 11). Our docking results also support the observation that DtpB preferentially binds hydrophobic peptides (Figure 5D), in agreement with previous data and an earlier study of the yeast homolog Ptr2p [70] and MD-based predictions for $PepT_{St}$ [71]. Interestingly, our analysis also reveals that tripeptide binders predominantly contain hydrophobic residues in position 2, whereas dipeptide binders may also contain hydrophilic residues in position 2 (Figure 5D).

Prediction and validation of peptide pose and binding for DtpB

To make use of the above described assays and tools, we hypothesize that the combination of computational docking prediction tools and experiments can be integrated in a workflow to accelerate the identification of DtpB substrates (Figure 6A). Firstly, the computed docking poses can be used to estimate the binder probability for all possible di- and tripeptides. Secondly, top hits are validated with thermal unfolding assays to determine the binding affinity, which can be performed experimentally on a medium to high-throughput level. Lastly, the peptides that exhibit intermediate binding affinity (see the bell-shaped dependence of transport rate vs K_D in Figure 4D) are analyzed for uptake by a low-throughput transport assay.

The obtained docking model (see DtpB preferentially binds small hydrophobic peptides) was developed with the input of twelve experimental structures of DtpB bound to various peptides determined in this study. The datasets of DtpB bound to dipeptides AL and NV have not been part of the training set and as such can be used to validate the predicted binding poses. RMSD values of the peptide backbone between the experimental structures and the top ten docking poses (rmsBB_offset see Materials & Methods) were 0.61 ± 0.16 Å for AL and 0.58 ± 0.15 Å for NV (Supplementary Figure 10). AUC of the modified hit selection procedure (when applied to all 200 generated dock poses) was 0.89 for AL and 0.9 for AV (Figure 6B), highlighting that the peptide binding pose inside the DtpB binding pocket can be predicted with high confidence. Furthermore, the RMSD for the peptide side-chains after superposition (rmsSC see Materials & Methods) was 1.45 ± 0.07 Å for AL

and 1.25 ± 0.24 Å for NV, suggesting that the conformation of the peptide can be predicted with good precision.

To test whether the classifier between binder and non-binder peptides (see DtpB preferentially binds small hydrophobic peptides) can correctly predict binders, we experimentally measured the binding affinity of twelve commercially available peptides that were not used to train the classifier (Supplementary Table 3). Of those, five peptides (AH, AY, GD, LLA, RGD) had binder probabilities above 0.5 (predicted binders) and seven peptides (EA, GPE, HH, PY, SH, YA, YYR) had binder probabilities in the range of 0 - 0.1 (predicted non-binders). Five out of seven non-binders were true negatives (experimental K_D over 1 mM), and two predicted non-binders (PY and SH) turned out to be false negatives with K_D values of 206 ± 126 and 352 ± 38 μM. In case of predicted binders, three out of five (AH, AY, LLA) were true positives (experimental K_D below 1 mM). Two of the false-positives among predicted binders (RGD and GD) were in the 'ambiguous' probability region (Figure 5B) with binder probabilities of only 0.52 and 0.6 (indicating low confidence of the prediction). The ROC AUC of the experimental validation of the classifier is 0.74, i.e. the discrimination capability of our classification approach is acceptable, yet there is room for improvement.

Finally, we asked whether selection of peptides using binder probability could exclude the peptides that are known to be transported (see Tight peptide binding is not associated with transport). As demonstrated in Figure 6C, only three out of twelve peptides transported by DtpB have low binder probability (predicted non-binders). Two peptides, including the reference substrate AA, are in the 'ambiguous' probability region, so they are unlikely to be discovered with the proposed workflow, however, the remaining seven known substrates can be potentially identified starting from the docking analysis.

Conclusions

In this work we present one of the most comprehensive structural and functional characterizations of a POT to date. We established a high-throughput crystallization pipeline, and determined the structures of 14 complexes of DtpB with ten different dipeptides and four tripeptides. Flexible residues within the binding site and multiple stabilizing polar interactions with peptides carrying various chemical groups were identified. Combined with a comprehensive biochemical characterization, these insights allowed us to quantify the binding probability of the whole di- and tripeptide space of proteinogenic amino acids and pin-point the key properties of a strong binder: high hydrophobicity and moderate size of the side-chain in the first position of the peptide.

A common assumption in biochemical studies of transporter proteins is that strong binders are also well transported [72]. For instance, it was recently demonstrated for three transporters from the MFS, amino acid-polyamine-organocation (APC) and mitochondrial carrier (MCS) superfamily, that stabilizing compounds identified by thermal unfolding assays are also well transported in follow-up radioactive measurements [72]. In case of DtpB, however, the trend is different, and only mid-affinity peptides (K_D between 100 µM and 2.5 mM) are well transported in a reconstituted system. We speculate that this could be a general feature of promiscuous transporters as opposed to highly specific transporters. We also note that transport assays are technically more demanding, so development of high-throughput and robust approaches would help advance our understanding of the mechanisms of transport and their relationship with binding. On the other hand, a combination of *in silico* predictions followed up by selection of potential interactors using a high-throughput assay can effectively narrow down the number of candidates to be tested in a low-throughput transport assay. In case of DtpB, applying such a funnel reduces the list of all possible di- and tripeptides (n = 8400) to a few hundreds of candidates that can be tested for binding by thermal unfolding assays or other techniques. Next, selected peptides in the medium affinity range can be tested for transport using liposome-based assays (pyranine assay or radioactive uptake measurement) ultimately identifying new peptide substrates for DtpB and potentially other promiscuous peptide transporters, including the clinically-relevant human PepT1 and PepT2 transporters. A critical step in this application would be to explore the moiety of the putative ligand that can mimic the N-terminus of a di- or tripeptide and be used for the initial placement of the ligand in the binding pocket. In this work we used a peptide-specific docking protocol, however, a scoring function for small molecules is available in Rosetta [73] thus significantly extending the scope of molecules that can be characterized *in silico*.

Our findings also indicate that high affinity peptides believed to be taken up by PepT1 in the human gut, could in fact act as inhibitors, justifying their use as attractive therapies in inflammatory bowel disease (IBD) and colonic cancer [74-79]. Overall, this work provides a solid molecular and biochemical basis for understanding how structural plasticity of POT's binding site allows for uptake of a large diversity of ligands. This constitutes a major step forward towards actual structure-based drug design approaches aiming at inhibiting these transport shuttle systems, or at hijacking them to increase drug absorption.

Acknowledgements

We thank the Sample Preparation and Characterization facility of EMBL Hamburg for support in this project and the beamlines P13 and P14 at EMBL Hamburg for regular access. We acknowledge Rolf

Nielsen for initial NanoDSF characterization of DtpB and Instruct-ERIC and the FWO for their support to the Nb discovery and Nele Buys for the technical assistance during nanobody discovery. Funding for the automated crystallography pipelines at EMBL Grenoble was provided by the grants iNEXT (g.n. 653706) and iNEXT Discovery (g.n. 871037) projects funded by the Horizon 2020 program of the European Commission and through Instruct-ERIC. All past and current members of the Löw group are acknowledged for their input to this manuscript. This research was supported through the Maxwell computational resources operated at Deutsches Elektronen-Synchrotron (DESY), Hamburg, Germany. VK and KJ were supported by a research fellowship from the EMBL Interdisciplinary Postdoc (EIPOD) Programme under Marie Curie Cofund Actions MSCA-COFUND-FP (grant agreement number 847543). The authors thank Ulrike Uhrig (EMBL ChemBio facility) for useful discussions on molecular docking.

Author contributions

Conceptualization: CL, MK, KJ, VK

Investigation: all authors

Software: VK

Writing – Original Draft: VK, CL, MK, KJ

Writing – Review & Editing: all authors

Visualization: VK, MK, KJ

Supervision: CL, MK, KJ, JAM, JS

Funding acquisition: CL, JAM, JS

Competing interests

The authors declare no competing financial interests

Data availability statement

The structures of the DtpB-peptide complexes were deposited to PDB with the following accession codes: 8B18, 8B19, 8B1A, 8B1B, 8B1C, 8B1D, 8B1E, 8B1F, 8B1G, 8B17, 8B1H, 8B1I, 8B1K, 8B1J. Data and code for the pyranine assay, thermal unfolding assay and docking-based prediction of

binder probability were deposited to Zenodo with record ID's 7612027, 7611944, 7612000, and 7586704.

Figures

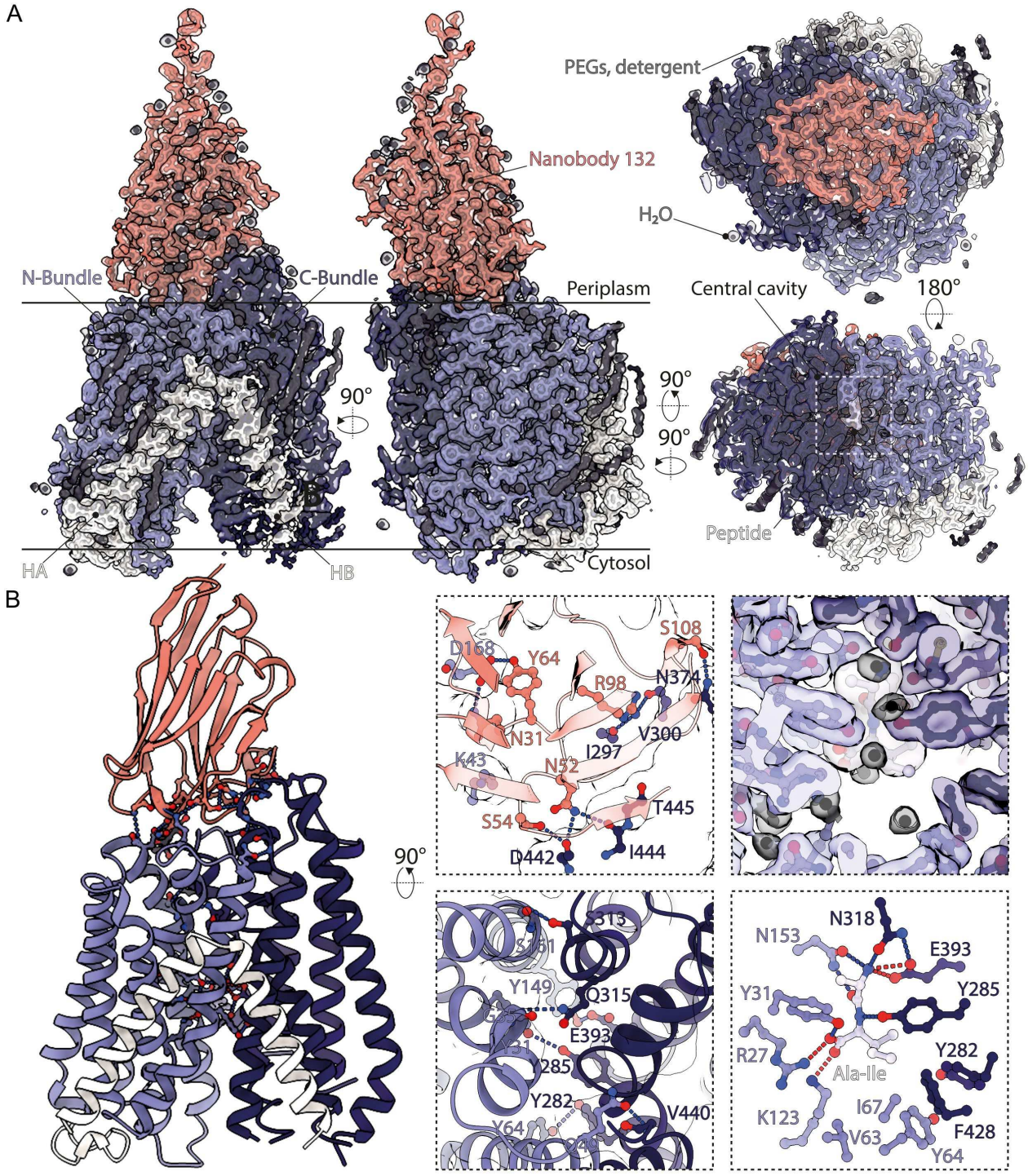


Figure 1: Structure of DtpB bound to nanobody Nb132 and the dipeptide AI. (A) The atomic model of DtpB-Nb132 fitted from the highest resolution dataset AI. The 2Fo-Fc map is shown as transparent surface (at σ = 1). The different structural elements are labeled. (B) Residues stabilizing the observed conformation are displayed as ball and sticks and the secondary structural elements are

473 shown as ribbons. Interactions between the transporter and the nanobody are shown in the top left
474 close up view, while interactions stabilizing the IF state between both bundles are highlighted in the
475 bottom panel. The electron density at the binding site and the dipeptide are illustrated in the top right
476 close up view, and the electrostatic interactions between the peptide and DtpB are shown as dashes
477 in the bottom panel (red dashed lines denote salt bridges; blue dashed lines correspond to polar
478 interactions; waters are shown in black).

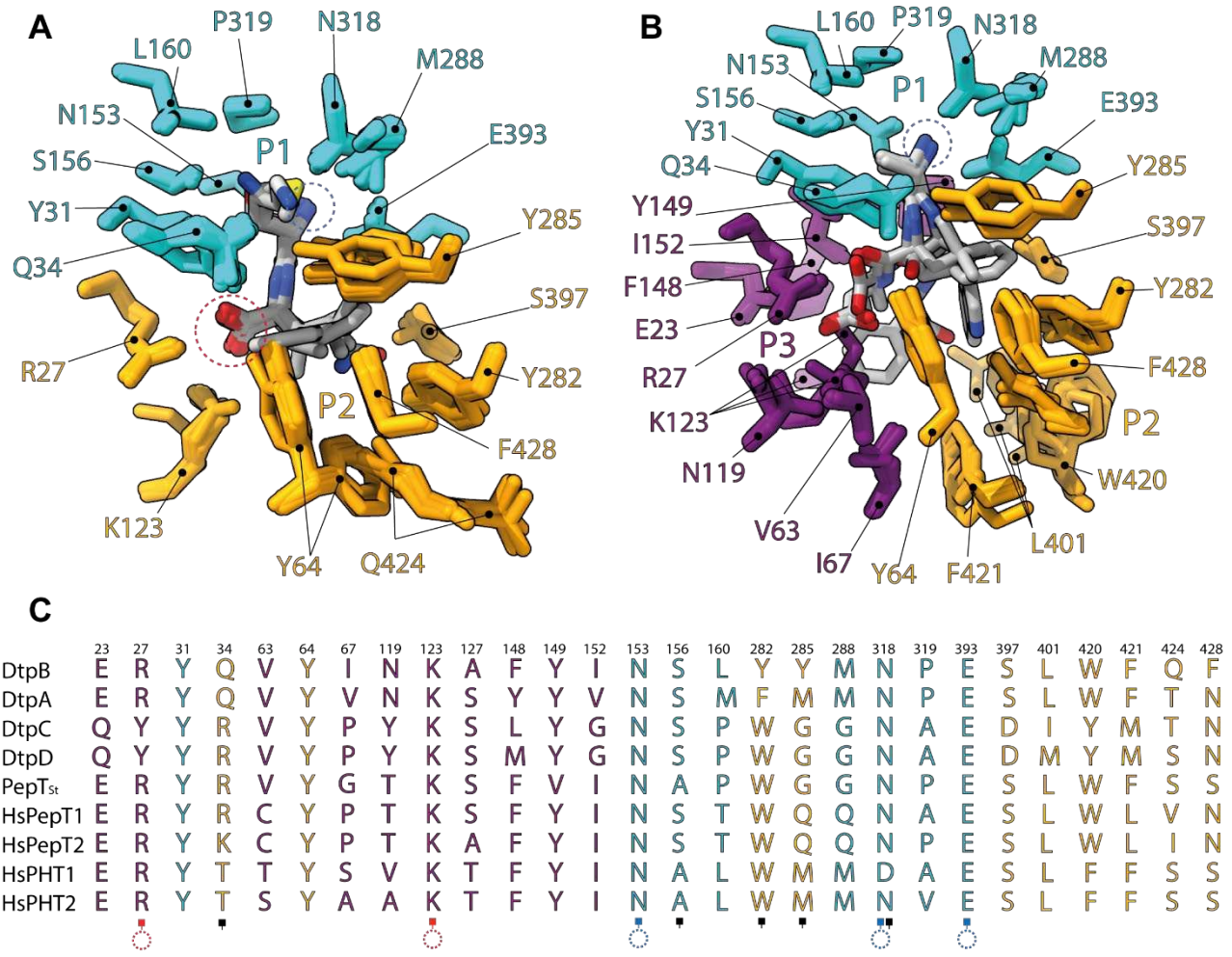


Figure 2: Definition of the binding pocket of DtpB. (A) Superimposition of the dipeptide co-crystal structures. (B) Superimposition of the tripeptide co-crystal structures. (C) Sequence alignment of the residues constituting the P1, P2, and P3 pockets in POTs. Residues constituting the P1, P2, and P3 pockets are respectively colored in cyan, yellow, and purple. Blue and red dashed rings circle the N-termini, and the C-termini. Blue and red squares indicate residues mediating electrostatic interactions with the termini of the co-crystallized peptides. Black squares indicate residues mediating polar interactions with the side chains of co-crystallized peptides. The N-termini are all coordinated in the same manner, while the C-termini adopt different positions in tripeptides.

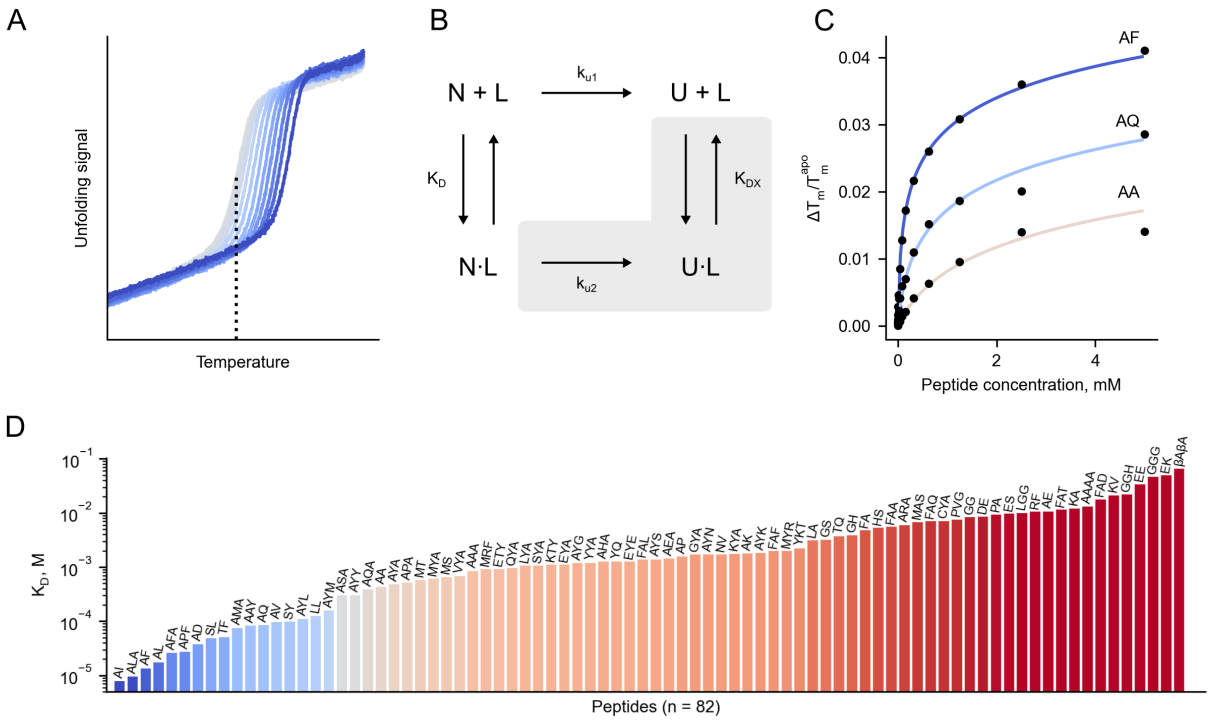


Figure 3: Large-scale determination of peptide binding to DtpB using a thermal unfolding assay. (A) Thermal unfolding assays can quantify protein-ligand interactions; ligand concentration increases with the blue color intensity. Melting temperature (T_m) of the apo state (gray curves) is shown with a vertical dashed line. (B) Hall's irreversible unfolding model to describe receptor-ligand interaction in a thermal unfolding assay. N - native state of the receptor, U - unfolded state of the receptor, L - ligand. Elements under the gray area denote the part of the model that was omitted in the current study. (C) Exemplary saturation curves of diverse peptides. Peptide affinity is color-coded with AF having the highest affinity, and AA having the lowest affinity. T_m^{apo} is the melting temperature of the apo state, ΔT_m is the difference in T_m between the apo state and in presence of respective peptide concentration. (D) Overview of K_D values obtained in this study; the affinity is color-coded with low-affinity peptides in red and high-affinity peptides in blue.

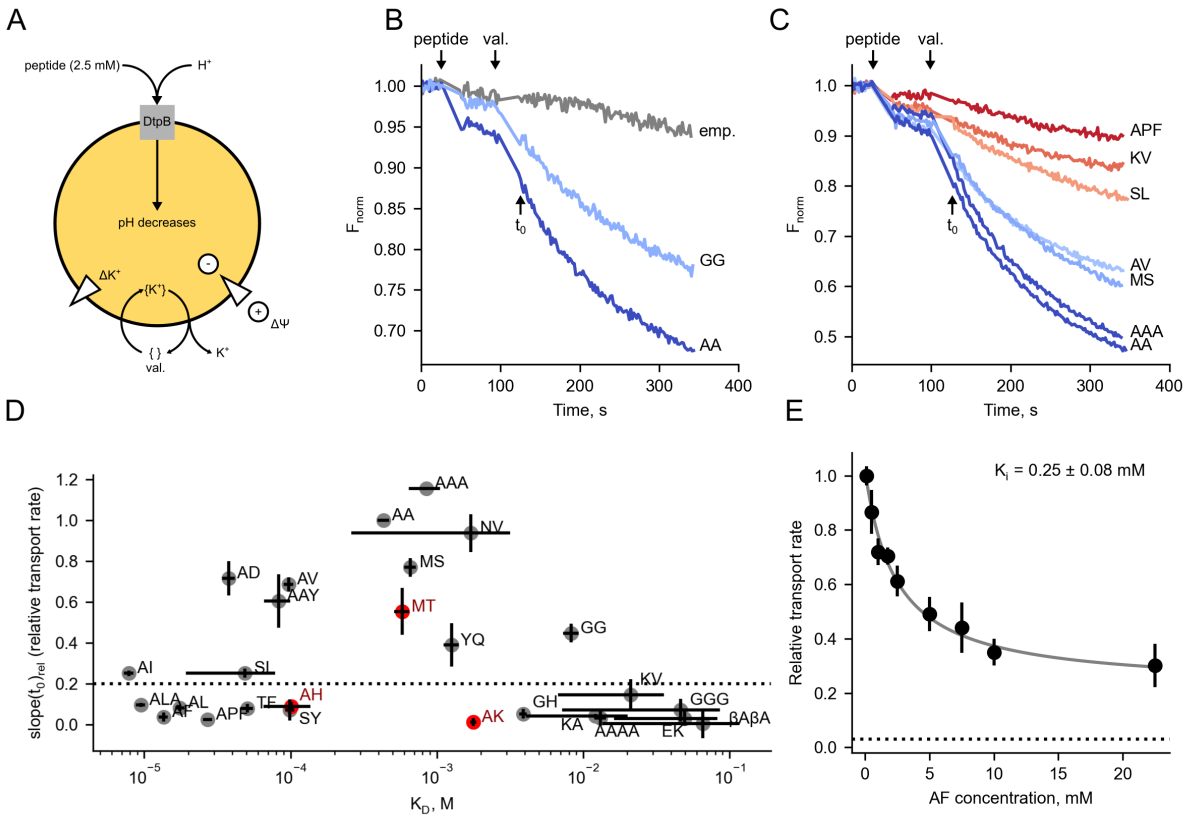
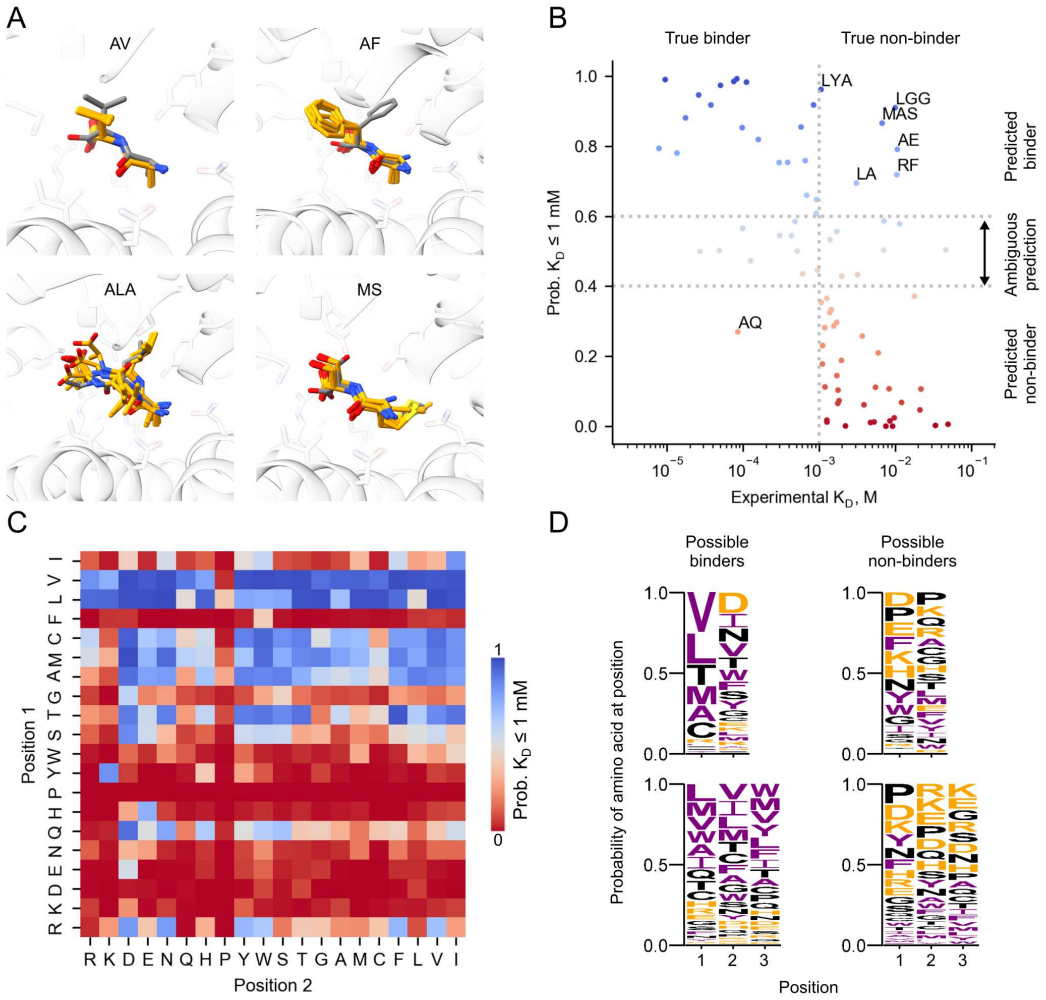


Figure 4: Measurement of peptide transport by DtpB and its relationship with peptide K_D. (A) The principle of the pyranine assay. DtpB is reconstituted into liposomes, and a concentration gradient of potassium ions (ΔK^+) from the inside to the outside of the liposome is created. Upon addition of valinomycin (val.) the potassium ions are chelated ($\{K^+\}$) and carried across the membrane and establish an electrochemical gradient (membrane potential $\Delta\Psi$). The membrane becomes hyperpolarized, and DtpB can utilize this gradient for proton-coupled peptide transport (peptide is added at indicated concentration to the outside buffer only). Proton flux into the liposome changes the fluorescence spectrum of the membrane-impermeable dye pyranine (yellow) present inside the liposome. (B) Exemplary transport curves of AA and GG detected with the pyranine assay. The transport signal from liposomes without DtpB (emp.) is shown in gray. Upper arrows denote approximate time when the peptide or valinomycin (val.) were added. Addition of peptide and valinomycin requires the fluorescence reading to be paused. Another arrow indicates t_0, which corresponds to the time point when the fluorescence readings are resumed. (C) Exemplary transport curves of diverse peptides tested in this work with pyranine assay. Approximate time when the peptide or valinomycin (val.) were added is shown with arrows. The curves are color-coded to show slow or no transport in red, and fast transport in blue. (D) Relationship between binding (K_D) and slope$(t_0)_{rel}$ (relative initial transport rate in pyranine assay). Dashed horizontal line indicates a cut-off of 0.20,

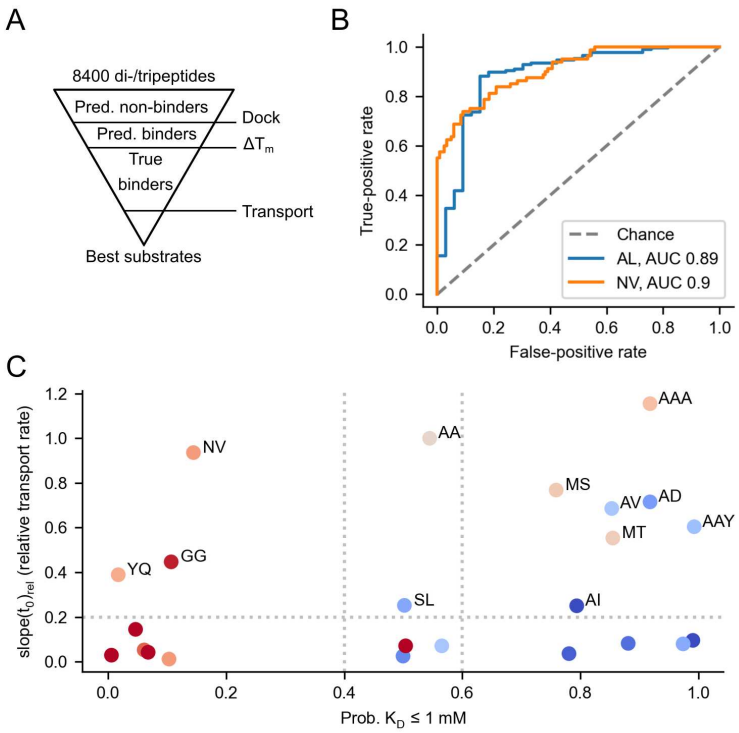
520 which corresponds to 20% of slope(t_0) of AA. Only peptides above the dashed line are considered to
521 be transported. Peptides marked in red were predicted to be transported based on their K_D and
522 subsequently tested in the pyranine assay. Error bars for K_D are standard deviations (n = 2). Error bars
523 for transport rate are the median absolute deviation (n = 2). (E) AF inhibits transport of AA. Relative
524 transport rates of AA in presence of variable AF concentrations are shown as black dots, and the fit
525 curve into Hill equation to determine IC_{50} is shown in grey. Error bars correspond to the standard
526 deviation (n = 3). The relative transport rate of AF alone is shown with a dashed line.
527



528

529 **Figure 5: Flexible docking of di- and tripeptides into DtpB.** (A) Exemplary docking results. The
530 binding pocket is viewed from the cytoplasmic side, and only the native structure of DtpB is shown.
531 The native conformation of the peptide is shown in gray, and top ten docked models are shown in
532 orange. For all 14 peptides with known experimental structures see SupFigure 10. (B)
533 Characterization of the logistic regression classifier that predicts binding of a peptide to DtpB.
534 Vertical dashed line shows the value of experimental K_D that is used to divide peptides into two
535 classes: binders ($K_D \leq 1$ mM) and non-binders ($K_D > 1$ mM). Horizontal dashed lines correspond to
536 the cut-off values for probabilities reported by the logistic regression classifier: if the binder
537 probability is below 0.4, then it is very likely that the peptide will not bind to DtpB. On the other
538 hand, with binder probabilities above 0.6 an interaction with DtpB is to be expected, however, it
539 cannot be excluded that the peptide will still exhibit low affinity. In the probability range between 0.4
540 and 0.6 it is difficult to assign a peptide to any class ('ambiguous' prediction). Data points are colored
541 by their binder probability (see color map in panel C). Labeled data points correspond to false-
542 positives (top right quadrant) and false-negatives (lower left quadrant). (C) Influence of the amino

543 acid identity on the probability of a dipeptide to be a DtpB binder. Amino acids in columns and rows
544 are ordered by Kyte-Doolittle hydropathicity scale [80] with most hydrophobic residues in the top right
545 corner. Note that in the Kyte-Doolittle scale W and Y are considered amphiphilic amino acids, so their
546 hydropathicity is close to 0. (D) Sequence logos (consensus sequence representation) for dipeptides
547 (top row) and tripeptides (bottom row) categorized by their probability to be a binder. Hydrophobic
548 residues are shown in purple, charged residues are colored orange.



549

550 **Figure 6: Characterization of the workflow to discover new peptide substrates for DtpB**. (A)
551 Schematic diagram of the proposed workflow. First, all possible di- and tripeptides (n=8400)
552 composed of proteinogenic amino acids are separated into predicted binders and non-binders based
553 on the docking results. This step is virtually instant, because the computation has already been
554 performed. Second, true binders are identified using high-throughput K_D estimation based on thermal
555 shifts (ΔT_m). Finally, selected peptides in the optimal K_D range for transport (between ~100 µM and
556 ~2.5 mM) are tested for transport using the low-throughput pyranine assay to identify true substrates.
557 (B) ROC for the prediction of peptide docking poses for AL and NV. (C) Peptides transported by
558 DtpB tend to have high binder probability. The shown data points are color-coded according to
559 affinity: high affinity (i.e. low K_D) in blue and low affinity (i.e. high K_D) in red. Only peptides above
560 the horizontal dashed line are considered transported. Vertical dashed lines indicate the 'ambiguous
561 probability' region, i.e. where the binder/non-binder nature of the peptide is predicted with low
562 confidence.

563

564

565

566 **Movie 1: Structural plasticity of the binding pocket of DtpB.** Structural overlay of the 14 different
567 DtpB-peptide complexes. Coordinating residues of the DtpB binding site are shown in sticks (light-
568 and dark-blue for residues of the N- and C-bundle). Peptides are illustrated in sticks and colored
569 white. Initially, a morph of ten DtpB structures bound to dipeptides is presented, but only the dipeptide
570 backbone (white) and coordinating backbone residues of the binding pocket are shown. This is
571 followed by a morph of the DtpB structures bound to tripeptides. Only minor structural changes on
572 the peptide backbone and coordinating residues of the transporter can be observed. This is followed
573 by a morph of all 14 structures and all residues of the binding pocket involved in peptide coordination
574 (including side-chains) are highlighted in light- and dark-blue sticks. The bound peptides are omitted
575 for clarity and the structural plasticity of the binding pocket is visualized. The binding pocket of DtpB
576 can adapt to the different side chains of the peptide, while the coordination of the peptide backbone
577 remains constant.

578

579

580

581

Tables

Table 1: Data collection and refinement statistics of DtpB-Nb132-peptide datasets

Peptide	AF	AFA	AI	AL	ALA	APF	AQ	AV	AW	AWA	KV	MS	NV	SL
PDB code	8B18	8B19	8B1A	8B1B	8B1C	8B1D	8B1E	8B1F	8B1G	8B17	8B1H	8B1I	8B1K	8B1J
Resolution (Å)	49.90 - 2.30 (2.38 - 2.30)	70.30 - 2.45 (2.55 - 2.45)	51.69 - 2.15 (2.23 - 2.15)	51.95 - 2.80 (2.90 - 2.80)	51.79 - 2.56 (2.65 - 2.56)	62.58 - 2.30 (2.38 - 2.30)	58.74 - 2.47 (2.56 - 2.47)	84.47 - 2.35 (2.43 - 2.35)	59.16 - 2.50 (2.59 - 2.50)	58.76 - 2.50 (2.59 - 2.50)	69.95 - 2.60 (2.69 - 2.60)	50.95 - 2.55 (2.64 - 2.55)	54.14 - 2.80 (2.90 - 2.80)	51.16 - 2.67 (2.77 - 2.67)
Space group	P 2 21 21	P 2 21 21	P 2 21 21	P 2 21 21	P 2 21 21	P 2 21 21	P 2 21 21	P 2 21 21	P 2 21 21	P 2 21 21	P 2 21 21	P 2 21 21	P 2 21 21	P 2 21 21
Cell dimensions a, b, c (Å)	54.42, 125.07, 168.71	54.70, 125.51, 169.72	54.32, 125.74, 168.00	54.59, 125.84, 169.03	54.40, 123.54, 169.33	54.76, 125.16, 169.90	54.47, 125.39, 168.18	54.65, 126.17, 168.94	54.61, 126.41, 168.04	54.60, 125.36, 168.82	54.34, 125.71, 168.37	54.18, 124.68, 167.46	54.14, 124.93, 165.77	54.49, 125.66, 168.04
α, ß, γ (º)	90, 9,0 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Total No. of reflections	667300 (64221)	578325 (55600)	810303 (71235)	378285 (38108)	75320 (7396)	105697 (10434)	83033 (8176)	649577 (45053)	81675 (7972)	80967 (8074)	72749 (7106)	75797 (7414)	57042 (5622)	65106 (5161)
No. of reflections	52149 (5144)	43863 (4296)	63584 (6245)	29527 (2894)	37673 (3699)	52854 (5217)	42145 (4129)	49598 (4863)	40878 (4000)	40497 (4037)	36376 (3553)	37899 (3707)	28523 (2811)	32661 (2628)
Completeness (%)	99.93 (99.94)	99.90 (99.95)	99.94 (99.94)	99.88 (100.00)	99.89 (99.78)	99.93 (99.96)	99.71 (99.61)	99.88 (99.94)	99.15 (98.62)	98.65 (100.00)	99.92 (100.00)	99.92 (99.95)	99.77 (99.93)	96.74 (77.81)
I/σI	16.79 (1.47)	16.45 (1.40)	18.87 (1.48)	16.30 (1.57)	16.74 (1.89)	13.91 (1.24)	14.56 (1.82)	15.18 (0.78)	12.36 (1.05)	14.29 (1.43)	13.06 (1.04)	9.44 (1.30)	12.01 (0.88)	12.66 (1.19)
Wilson B-factor (Å^2)	53.52	71.01	50.5	80.22	66.52	57.56	61.28	59.41	69.68	69.65	71.69	54.22	77.16	80.58
R_{merge}	0.08998 (1.649)	0.09384 (2.310)	0.07757 (1.664)	0.13730 (2.112)	0.02248 (0.4068)	0.02381 (0.5902)	0.02899 (0.4735)	0.11130 (3.0180)	0.01985 (0.6883)	0.02077 (0.4964)	0.02634 (0.6764)	0.04223 (0.5812)	0.03269 (0.8924)	0.02147 (0.675)
R_{meas}	0.09399 (1.719)	0.09789 (2.405)	0.08102 (1.742)	0.14350 (2.198)	0.03179 (0.5752)	0.03368 (0.8347)	0.0410 (0.6697)	0.11610 (3.199)	0.02807 (0.9734)	0.02937 (0.702)	0.03725 (0.9565)	0.05973 (0.822)	0.04623 (1.262)	0.03037 (0.9545)
R_{pim}	0.02673 (0.4834)	0.02738 (0.6643)	0.02302 (0.5110)	0.04087 (0.6025)	0.02248 (0.4068)	0.02381 (0.5902)	0.02899 (0.4735)	0.03238 (1.0420)	0.01985 (0.6883)	0.02077 (0.4964)	0.02634 (0.6764)	0.04223 (0.5812)	0.03269 (0.8924)	0.02147 (0.6750)
CC1/2	0.999 (0.648)	0.999 (0.522)	0.999 (0.591)	0.997 (0.531)	0.999 (0.694)	1 (0.609)	0.998 (0.640)	0.991 (0.285)	1 (0.432)	1 (0.643)	1 (0.488)	1 (0.545)	1 (0.367)	1 (0.350)
CC*	1 (0.887)	1 (0.828)	1 (0.862)	0.999 (0.833)	1 (0.905)	1 (0.870)	1 (0.883)	0.998 (0.666)	1 (0.777)	1 (0.885)	1 (0.810)	1 (0.840)	1 (0.733)	1 (0.720)
Reflections used in refinement	52127 (5142)	43856 (4295)	63564 (6244)	29508 (2894)	37667 (3699)	52830 (5216)	42133 (4122)	49579 (4862)	40846 (3993)	40479 (4037)	36360 (3553)	37882 (3705)	28489 (2809)	32590 (2577)
Reflections used for R_{free}	2589 (251)	2223 (236)	3113 (316)	2999 (306)	1892 (188)	2648 (249)	2078 (213)	2514 (252)	2015 (191)	2000 (193)	1807 (180)	1838 (200)	1382 (128)	1621 (122)
R_{work}	0.2267 (0.3774)	0.2241 (0.3115)	0.2154 (0.3490)	0.2254 (0.3354)	0.2151 (0.2944)	0.2299 (0.3515)	0.2221 (0.3108)	0.2164 (0.2998)	0.2245 (0.3442)	0.2314 (0.3054)	0.2257 (0.3659)	0.2241 (0.3319)	0.2260 (0.3869)	0.2211 (0.3589)
R_{free}	0.2536 (0.4101)	0.2673 (0.3354)	0.2390 (0.3854)	0.2633 (0.3708)	0.2410 (0.3293)	0.2575 (0.3866)	0.2669 (0.3498)	0.2363 (0.3244)	0.2481 (0.3548)	0.2647 (0.3278)	0.2602 (0.3766)	0.2574 (0.3686)	0.2503 (0.4129)	0.2636 (0.3964)
Number of non-hydrogen atoms	4718	4793	4787	4669	4781	4792	4786	4785	4786	4791	4787	4784	4789	4787
Protein residues	580	581	580	580	581	581	580	580	580	581	580	580	580	580
R.m.s. deviations Bond lengths (Å)	0.011	0.011	0.01	0.012	0.012	0.01	0.011	0.01	0.01	0.012	0.012	0.011	0.013	0.01
R.m.s. deviations Angles (º)	1.32	1.3	1.08	1.44	1.45	1.32	1.32	1.1	1.32	1.32	1.37	1.3	1.56	1.19
Ramachandran favored (%)	98.42	97.72	98.06	96.83	98.42	98.42	98.42	98.06	98.06	97.54	97.89	97.18	97.89	97.36
Ramachandran allowed (%)	1.41	2.11	1.94	3.17	1.58	1.58	1.58	1.94	1.94	2.46	2.11	2.82	1.94	2.64
Ramachandran outliers (%)	0.18	0.18	0	0	0	0	0	0	0	0	0	0	0.18	0
Rotamer outliers (%)	0.42	1.06	1.06	1.91	0.64	0.64	0.64	1.06	1.7	0.85	0.64	0.85	0.64	1.06
Clashscore	10.1	10.09	8.1	10.94	10.3	8.61	7.05	7.57	9.15	9.24	9.46	9.15	11.36	10.83
Average B-factor	67.63	80.49	60	82.94	75.22	67.08	64.85	67.4	77.25	75.84	75.99	61.75	82.31	81.74

Statistics for the highest-resolution shell are shown in parentheses.

References

1. Reitzer, L. (2003). Nitrogen assimilation and global regulation in Escherichia coli. Annu Rev Microbiol *57*, 155-176. 10.1146/annurev.micro.57.030502.090820.
2. Daniel, H., Spanier, B., Kottra, G., and Weitz, D. (2006). From bacteria to man: archaic proton-dependent peptide transporters at work. Physiology (Bethesda) *21*, 93-102. 10.1152/physiol.00054.2005.
3. Doki, S., Kato, H.E., Solcan, N., Iwaki, M., Koyama, M., Hattori, M., Iwase, N., Tsukazaki, T., Sugita, Y., Kandori, H., et al. (2013). Structural basis for dynamic mechanism of proton-coupled symport by the peptide transporter POT. Proc Natl Acad Sci U S A *110*, 11343-11348. 10.1073/pnas.1301079110.
4. Guettou, F., Quistgaard, E.M., Tresaugues, L., Moberg, P., Jegerschold, C., Zhu, L., Jong, A.J., Nordlund, P., and Low, C. (2013). Structural insights into substrate recognition in proton-dependent oligopeptide transporters. EMBO Rep *14*, 804-810. 10.1038/embor.2013.107.
5. Lyons, J.A., Parker, J.L., Solcan, N., Brinth, A., Li, D., Shah, S.T., Caffrey, M., and Newstead, S. (2014). Structural basis for polyspecificity in the POT family of proton-coupled oligopeptide transporters. EMBO Rep *15*, 886-893. 10.15252/embr.201338403.
6. Newstead, S., Drew, D., Cameron, A.D., Postis, V.L., Xia, X., Fowler, P.W., Ingram, J.C., Carpenter, E.P., Sansom, M.S., McPherson, M.J., et al. (2011). Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2. EMBO J *30*, 417-426. 10.1038/emboj.2010.309.
7. Ural-Blimke, Y., Flayhan, A., Strauss, J., Rantos, V., Bartels, K., Nielsen, R., Pardon, E., Steyaert, J., Kosinski, J., Quistgaard, E.M., and Low, C. (2019). Structure of Prototypic Peptide Transporter DtpA from E. coli in Complex with Valganciclovir Provides Insights into Drug Binding of Human PepT1. J Am Chem Soc *141*, 2404-2412. 10.1021/jacs.8b11343.
8. Zhao, Y., Mao, G., Liu, M., Zhang, L., Wang, X., and Zhang, X.C. (2014). Crystal structure of the E. coli peptide transporter YbgH. Structure *22*, 1152-1160. 10.1016/j.str.2014.06.008.
9. Jardetzky, O. (1966). Simple allosteric model for membrane pumps. Nature *211*, 969-970. 10.1038/211969a0.
10. Solcan, N., Kwok, J., Fowler, P.W., Cameron, A.D., Drew, D., Iwata, S., and Newstead, S. (2012). Alternating access mechanism in the POT family of oligopeptide transporters. EMBO J *31*, 3411-3421. 10.1038/emboj.2012.157.
11. Bartels, K., Lasitza-Male, T., Hofmann, H., and Low, C. (2021). Single-Molecule FRET of Membrane Transport Proteins. Chembiochem *22*, 2657-2671. 10.1002/cbic.202100106.
12. Drew, D., North, R.A., Nagarathinam, K., and Tanabe, M. (2021). Structures and General Transport Mechanisms by the Major Facilitator Superfamily (MFS). Chem Rev *121*, 5289-5335. 10.1021/acs.chemrev.0c00983.
13. Quistgaard, E.M., Low, C., Guettou, F., and Nordlund, P. (2016). Understanding transport by the major facilitator superfamily (MFS): structures pave the way. Nat Rev Mol Cell Biol *17*, 123-132. 10.1038/nrm.2015.25.
14. Newstead, S. (2017). Recent advances in understanding proton coupled peptide transport via the POT family. Curr Opin Struct Biol *45*, 17-24. 10.1016/j.sbi.2016.10.018.
15. Goh, E.B., Siino, D.F., and Igo, M.M. (2004). The Escherichia coli tppB (ydgR) gene represents a new class of OmpR-regulated genes. J Bacteriol *186*, 4019-4024. 10.1128/JB.186.12.4019-4024.2004.

16. Harder, D., Stolz, J., Casagrande, F., Obrdlik, P., Weitz, D., Fotiadis, D., and Daniel, H. (2008). DtpB (YhiP) and DtpA (TppB, YdgR) are prototypical proton-dependent peptide transporters of Escherichia coli. FEBS J *275*, 3290-3298. 10.1111/j.1742-4658.2008.06477.x.
17. Meredith, D., and Boyd, C.A. (2000). Structure and function of eukaryotic peptide transporters. Cell Mol Life Sci *57*, 754-778. 10.1007/s000180050040.
18. Killer, M., Finocchio, G., Mertens, H.D.T., Svergun, D.I., Pardon, E., Steyaert, J., and Low, C. (2022). Cryo-EM Structure of an Atypical Proton-Coupled Peptide Transporter: Di- and Tripeptide Permease C. Front Mol Biosci *9*, 917725. 10.3389/fmolb.2022.917725.
19. Chen, X.Z., Steel, A., and Hediger, M.A. (2000). Functional roles of histidine and tyrosine residues in the H(+)-peptide transporter PepT1. Biochem Biophys Res Commun *272*, 726-730. 10.1006/bbrc.2000.2851.
20. Foley, D.W., Rajamanickam, J., Bailey, P.D., and Meredith, D. (2010). Bioavailability through PepT1: the role of computer modelling in intelligent drug design. Curr Comput Aided Drug Des *6*, 68-78. 10.2174/157340910790980133.
21. Prabhala, B.K., Aduri, N.G., Iqbal, M., Rahman, M., Gajhede, M., Hansen, P.R., and Mirza, O. (2017). Several hPepT1-transported drugs are substrates of the Escherichia coli proton-coupled oligopeptide transporter YdgR. Res Microbiol *168*, 443-449. 10.1016/j.resmic.2017.01.005.
22. Prabhala, B.K., Aduri, N.G., Jensen, J.M., Ernst, H.A., Iram, N., Rahman, M., and Mirza, O. (2014). New insights into the substrate specificities of proton-coupled oligopeptide transporters from E. coli by a pH sensitive assay. FEBS Lett *588*, 560-565. 10.1016/j.febslet.2014.01.004.
23. Smith, M.W., Tyreman, D.R., Payne, G.M., Marshall, N.J., and Payne, J.W. (1999). Substrate specificity of the periplasmic dipeptide-binding protein from Escherichia coli: experimental basis for the design of peptide prodrugs. Microbiology (Reading) *145 (Pt 10)*, 2891-2901. 10.1099/00221287-145-10-2891.
24. Weitz, D., Harder, D., Casagrande, F., Fotiadis, D., Obrdlik, P., Kelety, B., and Daniel, H. (2007). Functional and structural characterization of a prokaryotic peptide transporter with features similar to mammalian PEPT1. J Biol Chem *282*, 2832-2839. 10.1074/jbc.M604866200.
25. Aduri, N.G., Prabhala, B.K., Ernst, H.A., Jorgensen, F.S., Olsen, L., and Mirza, O. (2015). Salt Bridge Swapping in the EXXERFXYY Motif of Proton-coupled Oligopeptide Transporters. J Biol Chem *290*, 29931-29940. 10.1074/jbc.M115.675603.
26. Ernst, H.A., Pham, A., Hald, H., Kastrup, J.S., Rahman, M., and Mirza, O. (2009). Ligand binding analyses of the putative peptide transporter YjdL from E. coli display a significant selectivity towards dipeptides. Biochem Biophys Res Commun *389*, 112-116. 10.1016/j.bbrc.2009.08.098.
27. Jensen, J.M., Aduri, N.G., Prabhala, B.K., Jahnsen, R., Franzyk, H., and Mirza, O. (2014). Critical role of a conserved transmembrane lysine in substrate recognition by the proton-coupled oligopeptide transporter YjdL. Int J Biochem Cell Biol *55*, 311-317. 10.1016/j.biocel.2014.09.016.
28. Jensen, J.M., Ernst, H., Wang, X., Hald, H., Ditta, A.C., Ismat, F., Rahman, M., and Mirza, O. (2012). Functional investigation of conserved membrane-embedded glutamate residues in the proton-coupled peptide transporter YjdL. Protein Pept Lett *19*, 282-287. 10.2174/092986612799363109.
29. Jensen, J.M., Ismat, F., Szakonyi, G., Rahman, M., and Mirza, O. (2012). Probing the putative active site of YjdL: an unusual proton-coupled oligopeptide transporter from E. coli. PLoS One *7*, e47780. 10.1371/journal.pone.0047780.

30. Jensen, J.M., Simonsen, F.C., Mastali, A., Hald, H., Lillebro, I., Diness, F., Olsen, L., and Mirza, O. (2012). Biophysical characterization of the proton-coupled oligopeptide transporter YjdL. Peptides *38*, 89-93. 10.1016/j.peptides.2012.08.012.
31. Killer, M., Wald, J., Pieprzyk, J., Marlovits, T.C., and Low, C. (2021). Structural snapshots of human PepT1 and PepT2 reveal mechanistic insights into substrate and drug transport across epithelial membranes. Sci Adv *7*, eabk3259. 10.1126/sciadv.abk3259.
32. Martinez Molledo, M., Quistgaard, E.M., Flayhan, A., Pieprzyk, J., and Low, C. (2018). Multispecific Substrate Recognition in a Proton-Dependent Oligopeptide Transporter. Structure *26*, 467-476 e464. 10.1016/j.str.2018.01.005.
33. Martinez Molledo, M., Quistgaard, E.M., and Low, C. (2018). Tripeptide binding in a proton-dependent oligopeptide transporter. FEBS Lett *592*, 3239-3247. 10.1002/1873-3468.13246.
34. Minhas, G.S., Bawdon, D., Herman, R., Rudden, M., Stone, A.P., James, A.G., Thomas, G.H., and Newstead, S. (2018). Structural basis of malodour precursor transport in the human axilla. Elife *7*. 10.7554/eLife.34995.
35. Nagamura, R., Fukuda, M., Kawamoto, A., Matoba, K., Dohmae, N., Ishitani, R., Takagi, J., and Nureki, O. (2019). Structural basis for oligomerization of the prokaryotic peptide transporter PepT(So2). Acta Crystallogr F Struct Biol Commun *75*, 348-358. 10.1107/S2053230X19003546.
36. Parker, J.L., Deme, J.C., Wu, Z., Kuteyi, G., Huo, J., Owens, R.J., Biggin, P.C., Lea, S.M., and Newstead, S. (2021). Cryo-EM structure of PepT2 reveals structural basis for proton-coupled peptide and prodrug transport in mammals. Sci Adv *7*. 10.1126/sciadv.abh3355.
37. Quistgaard, E.M., Martinez Molledo, M., and Low, C. (2017). Structure determination of a major facilitator peptide transporter: Inward facing PepTSt from Streptococcus thermophilus crystallized in space group P3121. PLoS One *12*, e0173126. 10.1371/journal.pone.0173126.
38. Shen, J., Hu, M., Fan, X., Ren, Z., Portioli, C., Yan, X., Rong, M., and Zhou, M. (2022). Extracellular domain of PepT1 interacts with TM1 to facilitate substrate transport. Structure *30*, 1035-1041 e1033. 10.1016/j.str.2022.04.011.
39. Stauffer, M., Jeckelmann, J.M., Ilgu, H., Ucurum, Z., Boggavarapu, R., and Fotiadis, D. (2022). Peptide transporter structure reveals binding and action mechanism of a potent PEPT1 and PEPT2 inhibitor. Commun Chem *5*, 23. 10.1038/s42004-022-00636-0.
40. Boggavarapu, R., Jeckelmann, J.M., Harder, D., Ucurum, Z., and Fotiadis, D. (2015). Role of electrostatic interactions for ligand recognition and specificity of peptide transporters. BMC Biol *13*, 58. 10.1186/s12915-015-0167-8.
41. Guettou, F., Quistgaard, E.M., Raba, M., Moberg, P., Low, C., and Nordlund, P. (2014). Selectivity mechanism of a bacterial homolog of the human drug-peptide transporters PepT1 and PepT2. Nat Struct Mol Biol *21*, 728-731. 10.1038/nsmb.2860.
42. Pardon, E., Laeremans, T., Triest, S., Rasmussen, S.G., Wohlkonig, A., Ruf, A., Muyldermans, S., Hol, W.G., Kobilka, B.K., and Steyaert, J. (2014). A general protocol for the generation of Nanobodies for structural biology. Nat Protoc *9*, 674-693. 10.1038/nprot.2014.039.
43. Cornaciu, I., Bourgeas, R., Hoffmann, G., Dupeux, F., Humm, A.S., Mariaule, V., Pica, A., Clavel, D., Seroul, G., Murphy, P., and Marquez, J.A. (2021). The Automated Crystallography Pipelines at the EMBL HTX Facility in Grenoble. J Vis Exp. 10.3791/62491.
44. Zander, U., Hoffmann, G., Cornaciu, I., Marquette, J.P., Papp, G., Landret, C., Seroul, G., Sinoir, J., Rower, M., Felisaz, F., et al. (2016). Automated harvesting and processing of protein crystals through laser photoablation. Acta Crystallogr D Struct Biol *72*, 454-466. 10.1107/S2059798316000954.

45. Biegel, A., Gebauer, S., Brandsch, M., Neubert, K., and Thondorf, I. (2006). Structural requirements for the substrates of the H+/peptide cotransporter PEPT2 determined by three-dimensional quantitative structure-activity relationship analysis. J Med Chem *49*, 4286-4296. 10.1021/jm0601811.
46. Brandsch, M. (2013). Drug transport via the intestinal peptide transporter PepT1. Curr Opin Pharmacol *13*, 881-887. 10.1016/j.coph.2013.08.004.
47. Boivin, S., Kozak, S., and Meijers, R. (2013). Optimization of protein purification and characterization using Thermofluor screens. Protein Expr Purif *91*, 192-206. 10.1016/j.pep.2013.08.002.
48. Chari, A., Haselbach, D., Kirves, J.M., Ohmer, J., Paknia, E., Fischer, N., Ganichkin, O., Moller, V., Frye, J.J., Petzold, G., et al. (2015). ProteoPlex: stability optimization of macromolecular complexes by sparse-matrix screening of chemical space. Nat Methods *12*, 859-865. 10.1038/nmeth.3493.
49. Pantoliano, M.W., Petrella, E.C., Kwasnoski, J.D., Lobanov, V.S., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B.A., Lane, P., and Salemme, F.R. (2001). High-density miniaturized thermal shift assays as a general strategy for drug discovery. J Biomol Screen *6*, 429-440. 10.1177/108705710100600609.
50. Vedadi, M., Arrowsmith, C.H., Allali-Hassani, A., Senisterra, G., and Wasney, G.A. (2010). Biophysical characterization of recombinant proteins: a key to higher structural genomics success. J Struct Biol *172*, 107-119. 10.1016/j.jsb.2010.05.005.
51. Bai, N., Roder, H., Dickson, A., and Karanicolas, J. (2019). Isothermal Analysis of ThermoFluor Data can readily provide Quantitative Binding Affinities. Sci Rep *9*, 2650. 10.1038/s41598-018-37072-x.
52. Layton, C.J., and Hellinga, H.W. (2010). Thermodynamic analysis of ligand-induced changes in protein thermal unfolding applied to high-throughput determination of ligand affinities with extrinsic fluorescent dyes. Biochemistry *49*, 10831-10841. 10.1021/bi101414z.
53. Lepock, J.R., Ritchie, K.P., Kolios, M.C., Rodahl, A.M., Heinz, K.A., and Kruuv, J. (1992). Influence of transition rates and scan rate on kinetic simulations of differential scanning calorimetry profiles of reversible and irreversible protein denaturation. Biochemistry *31*, 12706-12712. 10.1021/bi00165a023.
54. Sanchez-Ruiz, J.M. (1992). Theoretical analysis of Lumry-Eyring models in differential scanning calorimetry. Biophys J *61*, 921-935. 10.1016/S0006-3495(92)81899-4.
55. Hall, J. (2019). A simple model for determining affinity from irreversible thermal shifts. Protein Sci *28*, 1880-1887. 10.1002/pro.3701.
56. Jarmoskaite, I., AlSadhan, I., Vaidyanathan, P.P., and Herschlag, D. (2020). How to measure and evaluate binding affinities. Elife *9*. 10.7554/eLife.57264.
57. Fei, Y.J., Kanai, Y., Nussberger, S., Ganapathy, V., Leibach, F.H., Romero, M.F., Singh, S.K., Boron, W.F., and Hediger, M.A. (1994). Expression cloning of a mammalian proton-coupled oligopeptide transporter. Nature *368*, 563-566. 10.1038/368563a0.
58. Ganapathy, V., and Leibach, F.H. (1986). Carrier-mediated reabsorption of small peptides in renal proximal tubule. Am J Physiol *251*, F945-953. 10.1152/ajprenal.1986.251.6.F945.
59. Li, C., Yue, Z., Newstead, S., and Voth, G.A. (2022). Proton coupling and the multiscale kinetic mechanism of a peptide transporter. Biophys J *121*, 2266-2278. 10.1016/j.bpj.2022.05.029.
60. Kano, K., and Fendler, J.H. (1978). Pyranine as a sensitive pH probe for liposome interiors and surfaces. pH gradients across phospholipid vesicles. Biochim Biophys Acta *509*, 289-299. 10.1016/0005-2736(78)90048-2.
61. Parker, J.L., Mindell, J.A., and Newstead, S. (2014). Thermodynamic evidence for a dual transport mechanism in a POT peptide transporter. Elife *3*. 10.7554/eLife.04273.

62. Seifert, K., Fendler, K., and Bamberg, E. (1993). Charge transport by ion translocating membrane proteins on solid supported membranes. Biophys J *64*, 384-391. 10.1016/S0006-3495(93)81379-1.
63. Raveh, B., London, N., and Schueler-Furman, O. (2010). Sub-angstrom modeling of complexes between flexible peptides and globular proteins. Proteins *78*, 2029-2040. 10.1002/prot.22716.
64. Alford, R.F., Leaver-Fay, A., Jeliazkov, J.R., O'Meara, M.J., DiMaio, F.P., Park, H., Shapovalov, M.V., Renfrew, P.D., Mulligan, V.K., Kappel, K., et al. (2017). The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. J Chem Theory Comput *13*, 3031-3048. 10.1021/acs.jctc.7b00125.
65. Bender, B.J., Gahbauer, S., Luttens, A., Lyu, J., Webb, C.M., Stein, R.M., Fink, E.A., Balius, T.E., Carlsson, J., Irwin, J.J., and Shoichet, B.K. (2021). A practical guide to large-scale docking. Nat Protoc *16*, 4799-4832. 10.1038/s41596-021-00597-z.
66. Mobley, D.L., and Dill, K.A. (2009). Binding of small-molecule ligands to proteins: "what you see" is not always "what you get". Structure *17*, 489-498. 10.1016/j.str.2009.02.010.
67. Bishop, C.M. (2006). Probabilistic Discriminative Models. In Pattern Recognition and Machine Learning., p.209.
68. Klepsch, M.M., Kovermann, M., Low, C., Balbach, J., Permentier, H.P., Fusetti, F., de Gier, J.W., Slotboom, D.J., and Berntsson, R.P. (2011). Escherichia coli peptide binding protein OppA has a preference for positively charged peptides. J Mol Biol *414*, 75-85. 10.1016/j.jmb.2011.09.043.
69. Zainol, M.K.M., Linforth, R.J.C., Winzor, D.J., and Scott, D.J. (2021). Thermodynamics of semi-specific ligand recognition: the binding of dipeptides to the E.coli dipeptide binding protein DppA. Eur Biophys J *50*, 1103-1110. 10.1007/s00249-021-01572-y.
70. Ito, K., Hikida, A., Kawai, S., Lan, V.T., Motoyama, T., Kitagawa, S., Yoshikawa, Y., Kato, R., and Kawarasaki, Y. (2013). Analysing the substrate multispecificity of a proton-coupled oligopeptide transporter using a dipeptide library. Nat Commun *4*, 2502. 10.1038/ncomms3502.
71. Samsudin, F., Parker, J.L., Sansom, M.S.P., Newstead, S., and Fowler, P.W. (2016). Accurate Prediction of Ligand Affinities for a Proton-Dependent Oligopeptide Transporter. Cell Chem Biol *23*, 299-309. 10.1016/j.chembiol.2015.11.015.
72. Majd, H., King, M.S., Palmer, S.M., Smith, A.C., Elbourne, L.D., Paulsen, I.T., Sharples, D., Henderson, P.J., and Kunji, E.R. (2018). Screening of candidate substrates and coupling ions of transporters by thermostability shift assays. Elife *7*. 10.7554/eLife.38821.
73. Park, H., Zhou, G., Baek, M., Baker, D., and DiMaio, F. (2021). Force Field Optimization Guided by Small Molecule Crystal Lattice Data Enables Consistent Sub-Angstrom Protein-Ligand Docking. J Chem Theory Comput *17*, 2000-2010. 10.1021/acs.jctc.0c01184.
74. Dalmasso, G., Charrier-Hisamuddin, L., Nguyen, H.T., Yan, Y., Sitaraman, S., and Merlin, D. (2008). PepT1-mediated tripeptide KPV uptake reduces intestinal inflammation. Gastroenterology *134*, 166-178. 10.1053/j.gastro.2007.10.026.
75. Ingersoll, S.A., Ayyadurai, S., Charania, M.A., Laroui, H., Yan, Y., and Merlin, D. (2012). The role and pathophysiological relevance of membrane transporter PepT1 in intestinal inflammation and inflammatory bowel disease. Am J Physiol Gastrointest Liver Physiol *302*, G484-492. 10.1152/ajpgi.00477.2011.
76. Kovacs-Nolan, J., Zhang, H., Ibuki, M., Nakamori, T., Yoshiura, K., Turner, P.V., Matsui, T., and Mine, Y. (2012). The PepT1-transportable soy tripeptide VPY reduces intestinal inflammation. Biochim Biophys Acta *1820*, 1753-1763. 10.1016/j.bbagen.2012.07.007.
77. Miyake, M., Fujishima, M., and Nakai, D. (2017). Inhibitory Potency of Marketed Drugs for Ulcerative Colitis and Crohn's Disease on PEPT1. Biol Pharm Bull *40*, 1572-1575. 10.1248/bpb.b17-00181.

831 78. Viennois, E., Pujada, A., Zen, J., and Merlin, D. (2018). Function, Regulation, and
832 Pathophysiological Relevance of the POT Superfamily, Specifically PepT1 in Inflammatory
833 Bowel Disease. Compr Physiol *8*, 731-760. 10.1002/cphy.c170032.
834 79. Zhu, W., Ren, L., Zhang, L., Qiao, Q., Farooq, M.Z., and Xu, Q. (2020). The Potential of
835 Food Protein-Derived Bioactive Peptides against Chronic Intestinal Inflammation.
836 Mediators Inflamm *2020*, 6817156. 10.1155/2020/6817156.
837 80. Kyte, J., and Doolittle, R.F. (1982). A simple method for displaying the hydropathic
838 character of a protein. J Mol Biol *157*, 105-132. 10.1016/0022-2836(82)90515-0.
839